Football Match Result Prediction

1. Introduction

This project explores football match prediction using machine learning techniques. The dataset is publicly available on Kaggle: https://www.kaggle.com/datasets/technika148/football-database, which contains detailed match data from top European leagues.

Our goal is to predict the outcome of football games based on historical match statistics and advanced feature engineering. The target variable, gameresult, classifies each match as either a home win, draw, or away win.

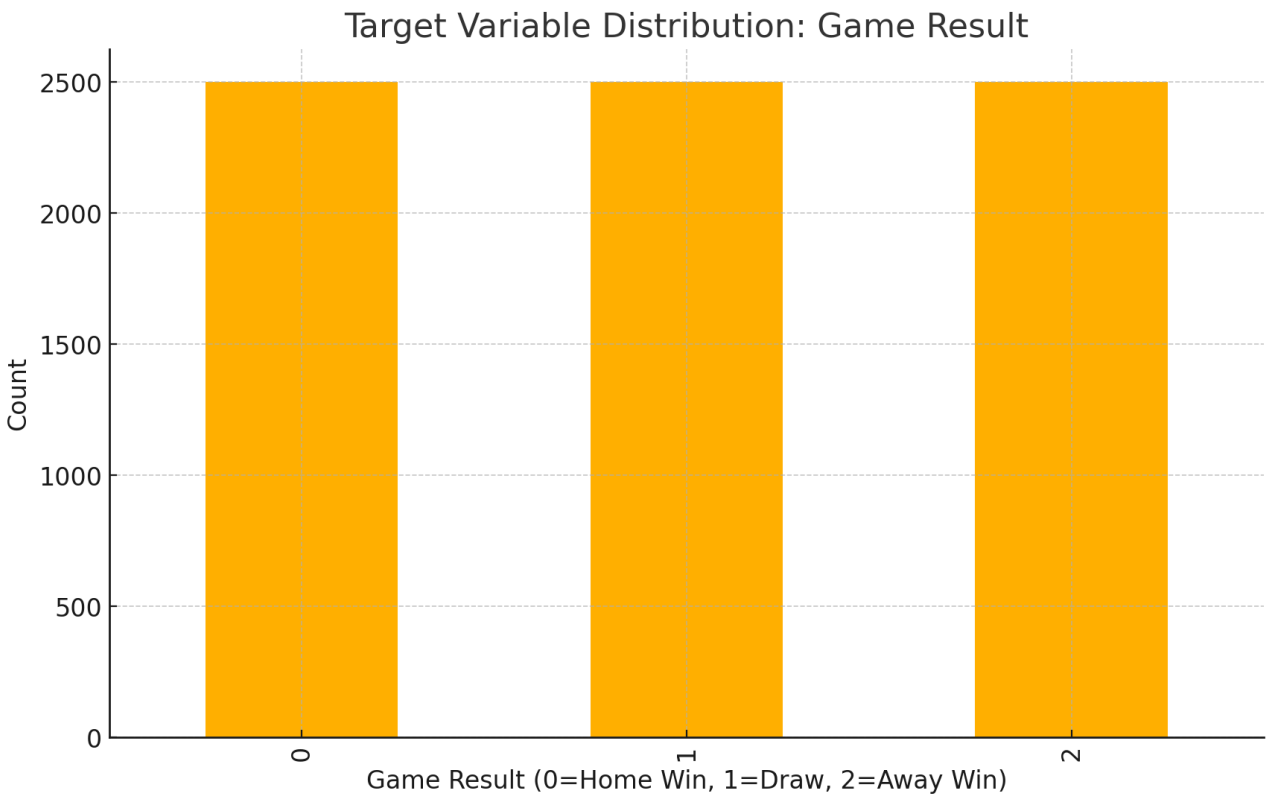
2. Objective

The main objective of this project is to build a classification model capable of predicting the result of football matches. It uses pre-match and in-game statistics to estimate the probability of each outcome (home win, draw, away win).

3. Dataset

The data includes match details from five top European leagues. Each row corresponds to a game, with over 100 features, including goals, assists, xGoals, fouls, corners, card counts, expected statistics, and more. The data is relatively balanced across the target classes.

Target Variable Distribution

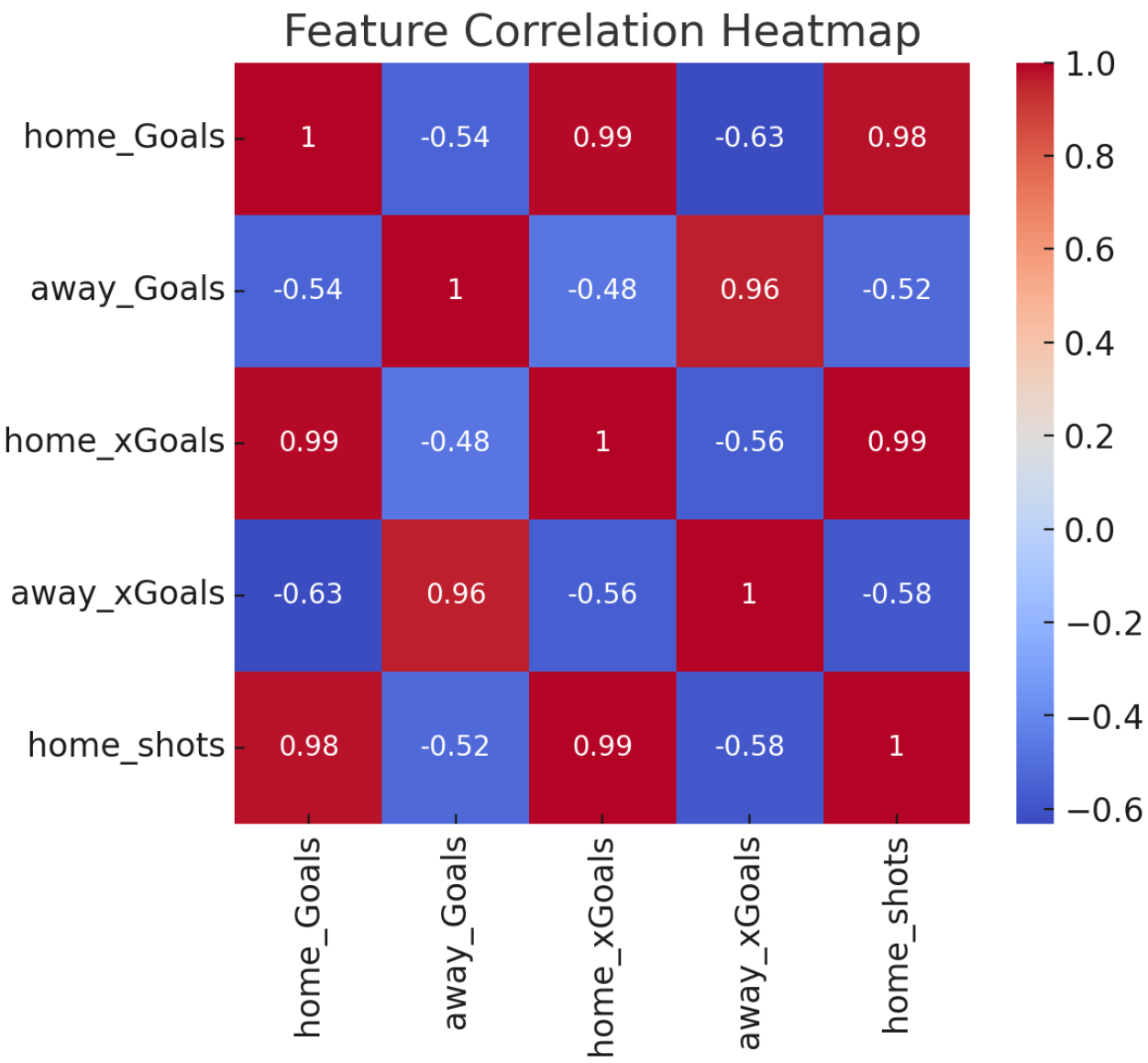


4. Project Design & Methodology

The methodology followed a full machine learning pipeline:

- Data Preprocessing
- Exploratory Data Analysis (AutoViz & Manual)
- Missing & Outlier Handling
- Feature Engineering
- Feature Selection
- Model Training & Evaluation
- Hyperparameter Tuning (Random Search & Grid Search)

Feature Correlation Heatmap



5. Models Used

The classification models applied were:

- Logistic Regression

- Support Vector Machine (SVM)

- Decision Tree

- Random Forest

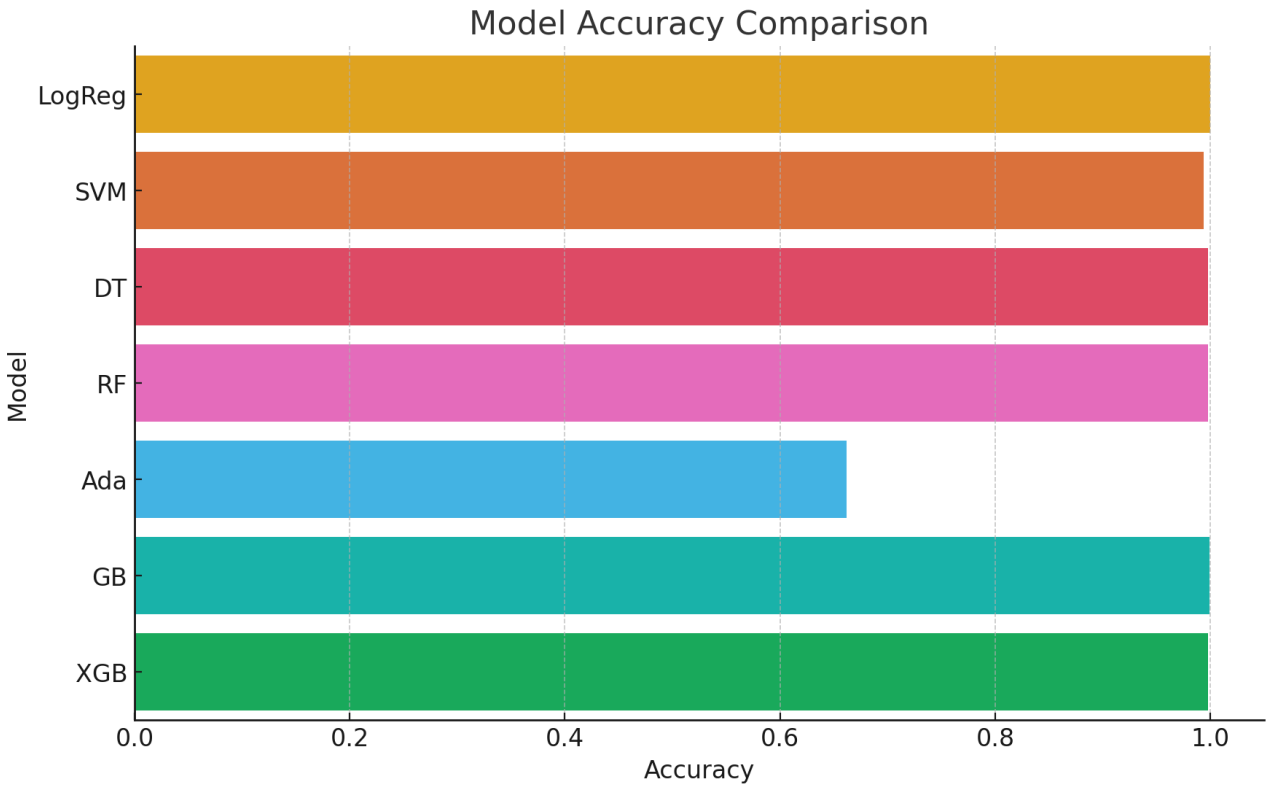
- AdaBoost

- Gradient Boosting

- XGBoost

Best performance was achieved using XGBoost, with over 99% accuracy.

Model Accuracy Comparison



6. Final Model Deployment

The final model selected for deployment:

```
XGBClassifier(
  learning_rate=0.05, max_depth=110, min_child_weight=50,
  subsample=0.8, n_estimators=400, objective='multi:softprob',
  eval_metric='logloss', random_state=0
)
```

It showed strong performance across all datasets and is ready for integration.

Football Match Result Prediction

Thank you

Leonardo Romano

Notebook Stages

1. Data Preprocessing

2. EDA (AutoViz)

3. EDA (Manual)

4. Outliers and Missing Values

5. Feature Engineering

6. Feature Selection

7. Classification Model and Hyperparameter Finetuning